

# Support Vector Machines

---

Nipun Batra

July 20, 2025

IIT Gandhinagar

# Non-Linearly Separable Data

pik1.png

Data not separate in  $\mathbb{R}$

Can we still use SVM?

Yes!

How? Project data to a higher dimensional space.

# Non-Linear Separable Data

pik2.png

# Non-Linearly Separable Data

pik3.png

## Another Example Transformation

pik4.png

# Projection/Transformation Function

$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$$

where,  $d$  = original dimension

$D$  = new dimension

In our example:

$$d = 1; D = 2$$

Linear SVM:

Maximize

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \bar{x}_i \cdot \bar{x}_j$$

such that constraints are satisfied.

↓

Transformation ( $\phi$ )

↓

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \phi(\bar{x}_i) \cdot \phi(\bar{x}_j)$$

## Trivial Example (Again)

pik5.png



# Steps

1. Compute  $\phi(x)$  for each point

$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$$

Q. If  $D \gg d$

Both steps are expensive!

1. Compute  $\phi(x)$  for each point

$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$$

2. Compute dot products over  $\mathbb{R}^D$  space

Q. If  $D \gg d$

Both steps are expensive!

Can we compute  $K(\bar{x}_i, \bar{x}_j)$

s.t.

$$K(\bar{x}_i, \bar{x}_j) = \phi(\bar{x}_i) \cdot \phi(\bar{x}_j)$$

where,

$K(\bar{x}_i, \bar{x}_j)$  is some function of dot product in original dimension

$\phi(\bar{x}_i) \cdot \phi(\bar{x}_j)$  is dot product in high dimensions (after transformation)

# Kernel Trick

pik6.png

$$\phi(x) = \langle \sqrt{2}x, x^2 \rangle$$

$K(x_i, x_j) = (1 + x_i x_j)^2 - 1$  where  $x_i x_j$  is dot product in lower dimensions

$$(1 + x_i x_j)^2 - 1 = 1 + 2x_i x_j + x_i^2 x_j^2 - 1$$

$$= \langle \sqrt{2}x_i, x_i^2 \rangle \cdot \langle \sqrt{2}x_j, x_j^2 \rangle$$

$$= \phi(x_i) \cdot \phi(x_j)$$

pik7.png

$\phi(x_1) = \langle -2\sqrt{2}, 4 \rangle; \phi(x_2) = \langle -\sqrt{2}, 1 \rangle$  Transformation

$\phi(x_1)\phi(x_2) = -2\sqrt{2} \times -\sqrt{2} + 4 \times 1 = 8$  Dot product in 2D

$K(x_1, x_2) = \{1 + (-2) \times (-1)\}^2 - 1$  Dot product in 1D

Q) Why did we use dual form?

Kernels again!!

Primal form doesn't allow for the kernel trick

$K(\bar{x}_1, \bar{x}_2)$  in dual and compute  $\phi(x)$  and then dot product in  $D$  dimensions

## Gram Matrix: (Positive Semi-Definite)

$$K(x_i, x_j) = (1 + x_i x_j)^2$$

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$
$x_1$	24	8	0	0	8	24	48
$x_2$	8	1	0	-1	0	...	
$x_3$	0	...	...	...	...	...	...
$x_4$	0						
$x_5$	8						
$x_6$	24						
$x_7$	48						

## Another Example

Q) What is  $\phi(x)$ ?

3. png

$$K(\bar{x}, \bar{x}') = \phi(\bar{x})\phi(\bar{x}')$$

$$K(\bar{x}, \bar{x}') = \left\{ \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix} \right\}^2 = (x_1 x'_1 + x_2 x'_2)^2$$

$$\Rightarrow \phi(x) = \langle x_1^2, \sqrt{2}x_1x_2, x_2^2 \rangle = x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x_1' x_2 x_2'$$



## Another Example

pik9.png

1. Linear:  $K(\bar{x}_1, \bar{x}_2) = \bar{x}_1 \bar{x}_2$

## Some Kernels

1. Linear:  $K(\bar{x}_1, \bar{x}_2) = \bar{x}_1 \bar{x}_2$
2. Polynomial:  $K(\bar{x}_1, \bar{x}_2) = (\rho + \bar{x}_1 \bar{x}_2)^q$

## Some Kernels

1. Linear:  $K(\bar{x}_1, \bar{x}_2) = \bar{x}_1 \bar{x}_2$
2. Polynomial:  $K(\bar{x}_1, \bar{x}_2) = (\rho + \bar{x}_1 \bar{x}_2)^q$
3. Gaussian:  $K(\bar{x}_1, \bar{x}_2) = e^{-\gamma \|\bar{x}_1 - \bar{x}_2\|^2}$  where  $\gamma = \frac{1}{2\sigma^2}$  - Also called Radial Basis Function (RBF)

Q) For  $\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  what space does kernel  $K(\bar{x}, \bar{x}') = (1 + \bar{x}\bar{x}')^3$  belong to?

$$\bar{x} \in \mathbb{R}^2$$

$$\phi(\bar{x}) \in \mathbb{R}^?$$

$$K(x, z) = (1 + x_1 z_1 + x_2 z_2)^3$$

$$= \dots$$

$$= \langle 1, x_1, x_2, x_1^2, x_2^2, x_1^2 x_2, x_1 x_2^2, x_1^3, x_2^3, x_1 x_2 \rangle$$

10 dimensional?

Q) For  $\bar{x} = x$ ; what space does RBF kernel lie in?

$$\begin{aligned}K(x, z) &= e^{-\gamma \|x-z\|^2} \\ &= e^{-\gamma (x-z)^2}\end{aligned}$$

Now:

$$e^{\alpha} = \sum_{n=0}^{\infty} \frac{\alpha^n}{n!}$$

$\therefore e^{-\gamma (x-z)^2}$  is  $\infty$  dimensional!!

Q) Is SVM parametric or non-parametric?

## SVM: Parametric or Non-Parametric

Q) Is SVM parametric or non-parametric?

Yes and No

Yes  $\rightarrow$  Linear kernel or polynomial kernel (form fixed)

No  $\rightarrow$  RBF (form changes with data)



## RBF is Non-Parametric

$$\begin{aligned}\hat{y}(x_{test}) &= \text{sign}(\bar{w}\bar{x}_{test} + b) \\ &= \text{sign}\left(\sum_{j=1}^{N_{SV}} \alpha_j y_j \bar{x}_j \bar{x}_{test} + b\right) \\ \hat{y}(X_{test}) &= \text{sign}\left(\sum_{j=1}^N \alpha_j y_j K(\bar{x}_j, \bar{x}_{test}) + b\right)\end{aligned}$$

$\alpha_j = 0$  where  $j \neq \text{S.V.}$

## RBF is Non-Parametric

Now  $K(\bar{x}_j, \bar{x}_{test})$  for RBF is:

$$e^{-\gamma ||\bar{x}_j - \bar{x}_{test}||^2}$$

∴ Hypothesis is a function of “all” train points

Closer  $\bar{x}$  is to  $\bar{x}_N$ ; more is it influencing  $\hat{y}(\bar{x})$  - hypothesis function

pik10.png

## RBF is Non-Parametric

- Now if we add a point to the dataset

## RBF is Non-Parametric

- Now if we add a point to the dataset
- Functional form can adapt (similar to KNN)

## RBF is Non-Parametric

- Now if we add a point to the dataset
- Functional form can adapt (similar to KNN)
- $\therefore$  SVM with RBF kernel is non-parametric

# Interpretation of RBF

- $\hat{y}(x) = \text{sign}(\sum \alpha_i y_i e^{-\|x - x_i\|^2} + b)$

# Interpretation of RBF

- $\hat{y}(x) = \text{sign}(\sum \alpha_i y_i e^{-||x-x_i||^2} + b)$
- $-||x - x_i||^2$  corresponds to radial term

# Interpretation of RBF

- $\hat{y}(x) = \text{sign}(\sum \alpha_i y_i e^{-||x-x_i||^2} + b)$
- $-||x - x_i||^2$  corresponds to radial term
- $\sum \alpha_i y_i$  is the activation component



# Interpretation of RBF

- $\hat{y}(x) = \text{sign}(\sum \alpha_i y_i e^{-\|x - x_i\|^2} + b)$
- $-\|x - x_i\|^2$  corresponds to radial term
- $\sum \alpha_i y_i$  is the activation component
- $e^{-\|x - x_i\|^2}$  is the basis component

## RBF: Effect of $\gamma$

$\gamma$ : How far is the influence of a single training sample

pik11.png